Andmehalduse juhised

Andmekirjelduse juhis
Lisa 1 andmekirjelduse standard

Mai 2022

Versioon 1.8

Dokumendi ajalugu

|  |  |  |  |
| --- | --- | --- | --- |
| **ver** | **muutuse sisu** | **autor** | **kuupäev** |
| 1.0 | Juhise aluseks on 2019-20 koostatud juhis „Eesti andmehalduse metoodikaprojekt. Andmekirjelduse juhis“. Selle üheks osaks on Lisa 2 Andmekirjelduse standard | Raivo Ruusalepp, Kuldar Aas, Siim Aben, Veiko Berendsen | aug 2020 |
| 1.1 | Versiooni 1.0 Lisa 2 Andmekirjelduse standard on tõstetud eraldi dokumendiks.  | Veiko Berendsen | märts 2021 |
| 1.7 | Sisse on viidud versioon 1.0 tagasiside ja RIHAKEse arendusel tekkinud vajadused, täpsustatud DCAT võrdlus. MKM, STAT, RIA tagasisideks edastamine.  | Kuldar Aas | aprill 2022 |
| 1.8 | Sisse viidud v1.7 tagasiside kommentaarid ja parandused. Edastamine andmete võrgustiku tagasisideks.  | Kuldar Aas | mai 2022 |
| *1.9* | *Märkuste sisseviimine* | *Veiko Berendsen* |  |
| *2.0* | *Lõppversioon, avalikustamine* | *Veiko Berendsen* |  |

Sisukord

[1 Sissejuhatus 3](#_Toc103002263)

[1.1 Eessõna 3](#_Toc103002264)

[1.2 Käsitlusala 3](#_Toc103002265)

[1.3 Andmekirjelduse standardi kasutusjuhud 4](#_Toc103002266)

[2 Andmekirjelduse elemendid 5](#_Toc103002267)

[2.1 Andmestik ja tema struktuursed osad 5](#_Toc103002268)

[2.1.1 Andmestiku kirjeldus 5](#_Toc103002269)

[2.1.2 Andmestiku levituse kirjeldus 12](#_Toc103002270)

[2.1.3 Andmebaasi tabeli kirjeldus 14](#_Toc103002271)

[2.1.4 Andmeelemendi kirjeldus 14](#_Toc103002272)

[2.2 Sõnastikud 17](#_Toc103002273)

[2.2.1 Andme- või ärisõnastiku kirjeldus 17](#_Toc103002274)

[2.2.2 Andmesõnastiku termini kirjeldus 18](#_Toc103002275)

[2.2.3 Ärisõnastiku termini kirjeldus 19](#_Toc103002276)

# Sissejuhatus

## Eessõna

Käesolev andmekirjelduse juhise lisa 1 esitab andmekirjelduse standardi. Standard esitab loetelu kohustuslikest ja soovituslikest kirjelduselementidest ja nende semantikast. Standardi rakendamine organisatsioonis tagab andmekirjelduste masinmõistetavuse, ehk võimaldab masinliidestuste tegemise nii organisatsioonide vahel kui riigi kesksete süsteemidega (RIHAKE, RIHA, avaandmete teabevärav).

Andmekirjelduse standardi väljatöötamisel on võimalikult palju ära kasutatud rahvusvahelisi parimaid praktikaid. Peamiseks rahvusvaheliseks aluseks on DCAT-AP (DCAT Application Profile for data portals in Europe)[[1]](#footnote-1), vähemal määral on kasutatud DDI (Data Documentation Initiative) ja OWL (W3C Web Ontology Language) komponente. Rahvusvaheliste praktikate kasutamine võimaldab edaspidi kirjeldusi vahetada ka rahvusvaheliste osapooltega ning vähendab standardi pikaajalise haldamise kulusid.

Andmekirjelduse standardi versioon 1.8 on tagasiside kogumiseks mõeldud mustand. Kõik kommentaarid standardi kohta on oodatud aadressile andmed@mkm.ee. Enne standardi juurutamist soovitame ühendust võtta MKMiga!

## Käsitlusala

Ülevaade andmekirjelduse komponentidest ja komponentide vahelistest seostest on toodud joonisel 1.



Joonis : Andmekirjelduse standardis kaetud andmekirjelduse komponendid

**Andmestiku kui terviku ülevaatlikuks kirjeldamiseks** esitab standard kaks kirjelduskomponenti – andmestik ja levitus. Andmestiku kirjeldus võimaldab anda üldise ülevaate andmestiku sisust ja ulatusest, andmestikuga seotud osapooltest ja teistest andmestikest ning esitab andmete piirdaatumid. Seeläbi tagab kirjeldus andmestiku leitavuse ja eristatavuse suuremas kataloogis (RIHA, avaandmete teabevärav, asutuse andmekataloog). Samuti on standardisse sisse toodud levituse *(DCAT: distribution)* mõiste ja kirjeldus. DCAT standardi järgi on levitus konkreetses vormingus väljavõte andmestikust, levituse kirjeldus võimaldab edasi anda täpsemat teavet andmestiku vormingu ja taaskasutamise võimaluste kohta.

Standardi teine loogiline osa on **andmestiku sisu detailne kirjeldus.** Standard ühtlustab relatsioonilise andmebaasina[[2]](#footnote-2) hallatava andmestiku tabelite ja väljade (andmeelementide) kirjelduse. Sealjuures on tabeli kirjeldus kasutusel grupeeriva tehnilise olemina, mis annab andmebaasi mudeli konteksti üksikutele andmeelementidele. Üksikute andmeelementide kirjeldus võimaldab lähemalt selgitada kirjeldatava andmeelemendi sisu ning anda edasi tehnilisi parameetreid (kasutatud loendid, andmetüübid jms). Seeläbi võimaldab andmeelementide kirjeldus täpselt aru saada andmete taaskasutamisega seotud piirangutest ja võimalustest.

Standardi kolmandaks osaks on **mõisteline kirjeldamine**. Selleks on standardisse sisse toodud ärisõnastiku ja andmesõnastiku olemid. Ärisõnastik on organisatsiooni (või organisatsioonide üleselt kokku lepitud) terminite kogum, mis on piisavalt täpne ja täielik organisatsiooni tegevuste ja andmete katmiseks. Ärisõnastikke võib asutusel olla üks või mitu, samuti kasutab ärisõnastik termineid valdkonna või üldistest märksõnastikest, kui viimased on olemas.

Ärisõnastiku terminid ja andmestiku andmeelemendid viiakse kokku andmesõnastikus, mis on seega ühendav lüli organisatsiooni mõistete maailma ja reaalselt juurutatud andmebaaside ja andmeelementide vahel. Andmekirjelduse standard võimaldab kirjeldada eraldi ärisõnastikku, andmesõnastikku ja mõlemat tüüpi sõnastiku termineid ning siduda viimaseid konkreetsete andmeelementidega. Sõnastike ja terminite ühtlustatud kirjeldusmudel võimaldab saada mõistelist ülevaadet erinevates andmestikes sisalduvatest andmetest, andmestike omanikel leida lihtsalt kattuvusi andmestike vahel ning andmete taaskasutajatel tuvastada efektiivsemalt just neile vajalikke andmeid.



Joonis : Mõistelise ja andmestiku kirjelduse koostamise loogika ja seosed.

Andmestiku ja selle komponentide (tabel, andmeelement) kirjeldusi tuleb üldjuhul luua eraldiseisvalt ärisõnastike haldamisest. Samas on andmesõnastiku koostamiseks (ehk ärisõnastiku terminite ja andmeelementide ühendamiseks) vajalik vähemalt esialgsel kujul ärisõnastiku ja andmeelementide kirjelduse olemasolu. Vastav kirjelduste koostamise loogika on visualiseeritud joonisel 2.

## Andmekirjelduse standardi kasutusjuhud

Standardi esimeseks peamiseks kasutusjuhuks on **andmekirjelduste vahetamine RIHAKE[[3]](#footnote-3) rakenduse ja RIHA süsteemi vahel**. 2022 – 2023 aasta jooksul juurutatakse standard mõlemas mainitud komponendis, tulemusena on organisatsioonidel võimalik kasutada RIHAKEse rakendust standardile vastava andmekirjelduse koostamiseks, sh automatiseerib RIHAKE suure osa andmeelementide ja tabelite kirjelduse koostamisest. RIHAKEses koostatud kirjeldusi on omakorda võimalik edastada ja avalikustada RIHA süsteemis.

Standardi teiseks peamiseks kasutusjuhuks on **organisatsioonis RIHAKEse väliselt loodud andmekirjelduste vahetamine RIHA süsteemiga**. Juhul kui organisatsioonis on juba kasutusel andmekataloog või võrreldav andmekirjelduse lahendus, võimaldab standardi juurutamine seal koostatud kirjeldusi eksportida, RIHAsse edastada ja seal avalikustada. Sarnaselt on võimalik ka RIHAs avalikustatud andmekirjelduste automaatne pärimine ja taaskasutus organisatsioonide süsteemides.

**NB!** Standardis on kajastatud mitmeid kirjelduselemente, mille loomine toimub automaatselt RIHAKEse rakenduse poolt. Sellised kirjelduselemendid on üldjuhul rakendatavad ainult esimese kasutusjuhu puhul, mitte organisatsioonide süsteemide ja RIHA andmevahetuse kontekstis.

**NB!** Mainitud kasutusjuhud ei eelda, et andmekirjeldust vahetatakse tervikuna. Pigem on ette näha vajadus üksikute kirjelduskomponentide (näiteks andmestiku kirjeldus, andmeelementide kirjeldused, sõnastik) vahetamiseks omaette olemitena. Seega pole ka käesolevas standardis eeldatud, et kõik kirjelduselemendid juurutatakse ühiselt, vaid pigem kirjelduskomponentide kaupa.

**NB!** Lisaks praegusele standardile on andmevahetuseks ette nähtud ka JSON vormingus „manifesti“ skeemi loomine. Manifesti täpne kuju, semantika ja süntaks on hetkel lõplikult kokku leppimata ja seda tehakse RIHA edasise arendamise käigus 2022 – 2023 aastal. **Organisatsioonidelt on oodatud arvamused ja seisukohad manifesti loomise põhimõtete ja ülesehituse osas!**

Lisaks eelnevalt mainitud kahele kasutusjuhule on standardi rakendamine võimalik ka teistes olukordades – näiteks andmestike kirjelduste ühtlustamisel avaandmete teabevärava ja RIHA vahel, või asutuste vahelises andmekirjelduste edastamises ja vahetamises.

# Andmekirjelduse elemendid

## Andmestik ja tema struktuursed osad

Andmestiku tervikkirjeldus on hierarhiline ning koosneb:

* Andmestiku kirjeldusest: suurema andmekogumi üldine kirjeldus, Võimaldab andmestikku leida ja tema sisust ülevaadet saada.
* Levituse kirjeldusest: andmestikust tehtud väljavõtte kirjeldus. Võimaldab levitust leida ja tema sisust esialgselt ülevaadet saada.
* Tabeli kirjeldusest: relatsioonilistes andmebaasides realiseeritud andmestike tabelite tehniline kirjeldus. Tabel on selle standardi mõttes grupeeriv olem, mis võimaldab omavahel siduda üksikute andmeelementide kirjeldused.
* Andmeelemendi kirjeldusest: üksiku andmeelemendi detailne kirjeldus. Võimaldab aru saada välja täpsest sisust ja tehnilisest ülesehitusest.

**Märkus:** Andmestiku ja selle levituste, tabelite ja andmeelementide kirjeldused ei ole praeguses standardis omavahel otseselt (näiteks identifikaatorite ja viidete abil) seotud. Selle asemel defineeritakse hierarhilised seosed (tabeli kuuluvus andmestikku, andmeelemendi kuuluvus tabelisse) läbi JSON vormingu hierarhilise struktuuri.

### Andmestiku kirjeldus

Andmestik on terviklikult identifitseeritav ja hallatav andmete kogum. Andmekogude puhul käsitletakse andmestikuna ühte andmekogu, samas on võimalik lisaks defineerida alamandmestikud (näiteks: Ehitisregistri andmestik, mis koosneb ehitise üldandmete ja ehitusjärelevalve alamandmestikest). Andmestiku kirjeldus põhineb suures osas DCAT-AP 2.1.0 standardil, lisatud on RIHAKEse, RIHA või avaandmete teabevärava jaoks vajalikke kirjelduselemente.

| # | Elemendi nimetus | Määratlus ja kasutamine | Kohustuslik / korduv | Näide | DCAT-AP viide | DCAT-APURI / vahemik |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | **pealkiri** | Andmekogu korral on andmestiku pealkirjaks selle pidamist reguleerivas õigusaktis toodud ametlik nimetus. Muu andmestiku puhul õigusaktis toodud täielik nimetus või praktikas kasutatav täielik nimetus. | 1..1 | Ehitisregister; Sadamaregister. | 4.4.1 Title | dct:title / rdfs:Literal |
| 2 | **kirjeldus** | Andmete sisuline lühikirjeldus ja andmetega seonduvate üldiste eesmärkide kirjeldus. Andmekogu kirjeldusena tuleb kajastada selle asutamise õigusaktis toodud andmestiku sisu kirjeldus ja andmete eesmärk.**Märkus:** Mitte segi ajada elemendiga „koostamiseesmärk“, mille abil kajastatakse (organisatsiooni) funktsioone ja tegevusi, mille käigus andmed tekivad. | 1..1 | Keskkonnaregister on loodusvarade, looduspärandi, keskkonnaseisundi ja keskkonnategurite andmeid sisaldav riigi põhiregister. Keskkonnaregistri eesmärk on koondada kogu keskkonnaandmestik ühte registrisse, seostades selle kaudu kõik keskkonnaandmed ajas ja ruumis ning anda neile õiguslik tähendus, tagades sellega andmestiku usaldatavuse nii rahvusvahelisel kui siseriiklikul tasandil. | 4.4.1 description | dct:description / rdfs:Literal |
| 3 | **lühinimetus** | Andmestikule viitamisel kasutatav üldlevinud lühinimetus. Kui andmestik on juba RIHAs kirjeldatud, kasutatakse RIHA lühinimetust. **Märkus**: Lühinimetuse mõiste on hetkel juba RIHAs kasutusele. RIHA edasise arendamise käigus (2022 – 2023) vaadatakse lühinimetusele kohanduvad reeglid üle ja vajadusel täpsustatakse, muuhulgas kaalutakse nõudeid lühinimetuse riigiülesele unikaalsusele.**Tagasiside küsimus:** Kas lühinimetuse kasutamine identifikaatorina on mõistlik või on eelistatud täielikult masinloetava (näiteks numbrilise või GUID vormingus) ühtlustatud identifikaatorite skeemi kasutusele võtmine RIHAKEses, RIHAs ja avaandmete teabeväravas? | 1..1 | ametipalk (Statistikaameti palkade andmestik);haudi (Muhu valla kalmistute andmekogu) | - | - |
| 4 | õiguslik alus | Vahetult andmestiku loomise ja haldamise aluseks oleva õigusakti nimetus ja viide. Võimalusel täidetakse õigusakti nimetus ja Riigi Teataja viide (URL). Kui õigusakt ei ole Riigi Teatajas avaldatud, antakse organisatsiooni kodulehel paikneva õigusakti viide (URL).**Märkus:** Kirjeldusse ei pea kandma kaudselt andmestiku aluseks olevaid õigusakte (näiteks Avaliku teabe seadus, Eesti Vabariigi põhiseadus) vaid ainult otseselt andmestiku aluseks olevaid seaduseid või määruseid.  | 0..n | Valga valla jäätmevaldajate registri põhimäärus https://www.riigiteataja.ee/akt/407062018011  | 4.4.3 is referenced by | dct:isReferencedBy / rdfs:Resource |
| 5 | **valdkond** | Andmestiku valdkond Eesti märksõnastiku (EMS, https://ems.elnet.ee/) alusel. | 1..n | haridus/algharidus | 4.4.2 theme / category | dcat:theme,subproperty ofdct:subject / skos:Concept |
| 6 | märksõna | EMSist võetud üks või mitu märksõna. Eelistada tuleb üldisi organisatsiooni tegevusi kirjeldavaid termineid. | 0..n | jäätmekäitlus; liikluskorraldus | 4.4.2 keyword/ tag | dcat:keyword / rdfs:Literal |
| 7 | märksõna URI | Märksõna viide URI vormingus. **Märkus:** RIHAKE täidab kirjelduselemendi peale märksõna valimist automaatselt EMS püsiviitega.  | 0..n | https://ems.elnet.ee/id/EMS016903 | - | - |
| 8 | tüüp | Kirjeldatava andmestiku tüüp Dublin Core Type Vocabulary[[4]](#footnote-4) loendi alusel.**Märkus:** Vaikimisi on kirjelduselemendi väärtuseks "dataset" ehk andmestik, võimalusel saab kasutada ka täpsemat väärtust, näiteks "text" või "picture".  | 0..1 | Dataset | 4.4.3 Type | dct:type / skos:Concept |
| 9 | keel | Andmestikus kasutatud keel(ed). Kui kirjeldus pole täidetud, eeldatakse vaikimisi eesti keele kasutamist. Andmevahetuses kasutatakse ISO 639-2 standardi alusel defineeritud kolmetähelisi keele koode (https://www.loc.gov/standards/iso639-2/php/code\_list.php). **Märkus:** RIHAKEse kasutajaliides võimaldab keelte valimist inimloetavana, ISO 639-2 vormingusse teisendamine toimub rakenduses automaatselt.  | 0..n | est, ukr  | 4.4.3 language | dct:language / dct:LinguisticSystem |
| 10 | **omanik** | Andmestiku teabevaldaja (organisatsiooni) täielik nimetus. Andmekogu kirjeldamisel märgitakse vastutava töötleja nimetus, muude andmestike korral selle organisatsiooni nimetus, kelle valitsemise all andmed on.**Märkus:** RIHAKE täidab andmestiku omaniku kirjelduse automaatselt kirjeldajaga seotud organisatsiooni nimetusega.  | 1..1 | Majandus- ja kommunikatsiooniministeerium;Kihnu vallavalitsus | 4.4.2publisher | dct:Publisher / foaf:Agent |
| 11 | **haldaja** | Andmestikku haldava organisatsiooni täielik nimetus. Andmekogu puhul on haldajaks andmekogu volitatud töötleja, muude andmestike korral sisuliste tegevuste mõttes volitatud töötleja rolli täitev organisatsioon.**Märkus:** Andmestiku omanik ja haldaja võivad olla üks organisatsioon.  | 1..n | Registrite ja Infosüsteemide Keskus; Audru vallavalitsus | - | - |
| 12 | kontaktisik | Andmestiku kontaktisikuks määratud organisatsiooni töötaja nimi, amet, e-post ja telefon. Märkus: Kasutatakse vCard ontoloogiat, millest omakorda on kasutusel neli alamelementi (nimi, ametikoht, e-posti aadress, telefoninumber) | 0..n | Jüri Tamm, andmehaldur, jyri.tamm@mmit.ee, 6789012 | 4.4.2 contact point | dcat:contactPoint / vcard:Kind |
| 12.1 | kontaktisiku nimi | Kontaktisiku täisnimi. Igas kontaktisiku kirjes peab sisalduma täpselt üks nimi | 1..1[[5]](#footnote-5) | Jüri Tamm  | - | vcard:fn / rdfs:Literal |
| 12.2 | kontaktisiku ametikoht | Kontaktisiku ametikoht. Igas kontaktisiku kirjes võib sisalduda üks ametikoht | 0..1 | andmehaldur | - | vcard:hasRole / rdfs:Literal |
| 12.3 | kontaktisiku e-posti aadress | Kontaktisiku e-posti aadress. Igas kontaktisiku kirjes peab sisalduma vähemalt üks kahest, kas e-posti aadress või telefoninumber. | 0..1 | jyri.tamm@mmit.ee  | - | vcard:hasEmail / rdfs:Literal |
| 12.4 | kontaktisiku telefoninumber | Kontaktisiku telefoninumber. Igas kontaktisiku kirjes peab sisalduma vähemalt üks kahest, kas e-posti aadress või telefoninumber. | 0..1 | 6789012 | - | vcard:hasTelephone / rdfs:Literal |
| 13 | lauter | Andmestiku avalik veebiliides või muu veebisait millelt on võimalik andmestikku kasutada või sellele juurdepääsu taotleda.  | 0..1 | http://register.keskkonnainfo.ee/envreg/https://haridus.saaremaavald.ee/huviharidus/  | 4.4.3 landing page | dcat:landingPage / foaf:Document |
| 14 | koostamise eesmärk | Loetelu organisatsiooni tegevustest, mille käigus andmestik tekib ja täieneb. Andmekogu korral tuleks sisestada selle asutamise õigusaktis toodud tegevused. Muude andmestike puhul (või kui õigusakt tegevusi ei kajasta), organisatsiooni liigitusskeemis toodud sobivad funktsioonid või sarjade aluseks olevad tegevused. **Märkus:** Mitte segi ajada elemendiga „Kirjeldus“, milles esitatakse ülevaade andmestiku sisust (kogutavatest andmetest) ja andmekogumise eesmärkidest.**Tagasiside küsimus:** Kas selle elemendi täitmine peaks olema juurutatud riigi funktsioonide ja tegevuste sõnastiku abil? Kas olete valmis selle sõnastiku loomisesse panustama? | 0..n | Lemmiklooma registreerimine;Hauaplatsi hooldaja määramine | 4.4.2 was generated by | prov:wasGeneratedBy / prov:Activity |
| 15 | andmete päritolu | Teise andmekogu nimetus, millest andmestiku andmed on päritud. Võimalusel tuleb kasutada teise andmekogu (RIHA) lühinimetust. | 0..n | ehr; pjis-sf. | 4.4.2 source | dct:source / dcat:Dataset |
| 16 | eelkäija | Teise andmekogu nimetus, mis on olnud kirjeldatava andmestiku eelkäija.Võimalusel tuleb viitamisel kasutada teise andmekogu RIHA lühinimetust.**Tagasiside küsimus:** Kas eelkäija kirjeldamine andmestiku kirjelduse osana on mõistlik ja vajalik, või peaks kirjeldatava andmestiku päritolu ja eelkäija katma ühe kirjelduselemendiga? | 0..n | Hooneregister;skais; | 4.4.3 provenance | dct:provenance / dct:ProvenanceStatement |
| 17 | **kasutusele võtmise kuupäev** | Andmestiku pidamise algusaeg, kattub enamasti andmestiku infosüsteemi kasutusele võtmise ajaga. Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP)**Märkus:** Mitte segi ajada elemendiga „andmete piirdaatumid“ | 1..1 | 2003-01-10 | 4.4.3 release date | dct:issued / rdfs:Literal typed as xsd:date, xsd:dateTime, xsd:gYear or xsd:gYearMonth |
| 18 | kirjelduse muutmiskuupäev | Andmestiku kirjelduse viimase muutmise kuupäev.Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP) | 0..1 | 2020-05-25 | - | - |
| 19 | andmete piirdaatumid | Andmestikus sisalduvate andmete piirdaatumid, näitab millist ajaperioodi andmed katavad. Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP/AAAA-KK-PP)**Märkus:** Andmete piirdaatumeid ei tohi segi ajada andmekogumise või -sisestamise piirdaatumitega, samuti andmestiku kasutusele võtmise kuupäevaga. **Märkus:** Lõppdaatumi märkimine on tarvilik ainult suletud andmestike kirjeldamisel. | 0..1 | 1993-11-05/2007-06-301920/2000-022005-11//1940  | 4.4.2 temporal coverage | dct:temporal / dct:PeriodOfTime |
| 20 | andmete uuendamise regulaarsus | Andmestiku andmete uuendamise regulaarsus. Täita juhul kui andmete kogumine või loomine ei toimu pidevalt vaid perioodiliselt.Kirjelduselemendi väärtused võetakse EL sageduste sõnastikust (EU Frequency Vocabulary)[[6]](#footnote-6). | 0..1 | Kord aastas | 4.4.3 frequency | dct:accrualPeriodicity / dct:Frequency |
| 21 | andmete ruumiline ulatus | Andmete poolt kaetud geograafilise piirkonna nimetus. Eesti kohta kasutatakse jaotust vald/linn, maakond või kogu Eesti. Muu maailma kohta riigi nimetus, regiooni nimetus või globaalne. | 0..1 | Eesti;Mulgi vald;Põhjamaad;Soome | 4.4.2 spatial / geographical coverage | dct:spatial / dct:Location |

### Andmestiku levituse kirjeldus

Andmestiku levitus on andmestiku konkreetseks kasutusjuhuks mõeldud väljavõte või kasutusmeetod. Levituse kirjeldus võimaldab edasi anda täpsemat teavet andmestiku vormingu ja taaskasutamise võimaluste kohta. Levituse kirjeldamisel lähtutakse DCAT-AP 2.1.0 standardist.

Märkus: RIHAKEse, RIHA ja avaandmete teabevärava juurutustes ei eristata andmestiku ja levituse kirjeldust vaid on liidetud kokku ühtseks andmestiku kirjelduseks.

**Märkus:** DCAT-AP levituse kirjelduselemendid on peamiselt rakendatud avaandmete teabeväravas, v.a. kirjelduselement „kasutuslitsents“, mida kasutatakse ka RIHAKEse andmestiku kirjelduses.

| # | Elemendi nimetus | Määratlus ja kasutamine | Kohustuslik / korduv | Näide | DCAT-AP viide | DCAT-AP URI / vahemik |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | kasutuslitsents | Kogu andmestikule kohalduv kasutuslitsents, kui see on määratud. Elemendi täitmisel kasutakse Creative Commons litsentse. | 0..1 | CC BY-SA 4.0 | 4.5.2 licence | dct:license / dct:LicenseDocument |
| 2 | avalikustamise kestus | Kuupäev, millal levitus viimast päeva kättesaadav on. Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP)**Märkus:** Elementi ei kasutata RIHAKEses | 0..1 | 2025-06-31 | 4.5.2 availability | dcatap:availability / skos:Concept |
| 3 | andmete staatus | Element viitab levituse küpsusastmele. Avaandmete teabevärav eristab nelja staatust: valmis, ebasoovitav, arenduses, mahavõetud.**Märkus:** Elementi ei kasutata RIHAKEses | 0..1 | valmisebasoovitavarendusesmahavõetud | 4.5.3 status | adms:status / skos:Concept |
| 4 | juurdepääs | Levituse juurdepääsetavuse lühikirjeldus. Avaandmete teabevärav eristab kolme juurdepääsu tasandit: avalik, piiratud, suletud. **Märkus:** Elementi ei kasutata RIHAKEses | 0..1 | avalikpiiratudsuletud | 4.5.3 Rights | dct:rights / dct:RightsStatement |
| 5 | viide  | Otsene viide portaalis või muus andmestikke avaldavas ressursis olevale levitusele.**Märkus:** Elementi ei kasutata RIHAKEses | 0..1 | https://andmed.stat.ee/et/stat/majandus\_\_infotehnoloogia\_\_infotehnoloogia-leibkonnas/IT621 | 4.5.1 access URL | dcat:accessURL / rdfs:Resource |
| 6 | ajaline täpsus | Element viitab levituses sisalduvates andmetes minimaalsele eristatavale ajaperioodile.**Märkus:** Elementi ei kasutata RIHAKEses | 0..1 |  | 4.5.3 temporal resolution | dcat:temporalResolution / xsd:duration |

### Andmebaasi tabeli kirjeldus

Andmebaasi tabel on andmekirjelduse mudelis grupeeriv olem, mis seob kirjeldatava relatsioonilise andmebaasi andmeelemendid.

**Märkus:** RIHAKEse rakendus täidab tabeli kirjelduselemendid automaatselt andmebaasi skanneerimise käigus.

| # | Elemendi nimetus | Määratlus ja kasutamine | Kohustuslik / korduv | Näide |
| --- | --- | --- | --- | --- |
| 1 | **nimetus** | Tabeli nimetus täpselt sellisel kujul, nagu see on esitatud relatsioonilises andmebaasis.  | 1..1 | SYS\_LOGEHITISISIK |
| 2 | schema | Andmebaasi skeemi (schema) nimetus, millesse tabel kuulub. | 0..1 | SYSRIHARIHA\_vana |
| 3 | **andmehoidla** | Andmehoidla (andmebaasi) nimetus, millesse tabel kuulub. | 1..1 | rihaehrrr |
| 4 | kommentaar | Tabeli kommentaar andmebaasis | 0..1 | Ehitise üldandmete tabel |
| 4 | kirjelduse muutmise aeg | Tabeli kirjelduse viimase muutmise aeg**Märkus:** RIHAKEses kattub viimase muutmise aeg andmebaasi viimase skanneerimise ajaga.  | 0..1 | 2022-03-17T10:10:34,9344 |

### Andmeelemendi kirjeldus

Andmeelemendi kirjeldus esitab üksiku andmeelemendi detailse kirjelduse, mille abil on võimalik aru saada välja sisust, tähendusest ja seostest teiste andmeelementidega. Standardis toodud andmeelemendi kirjeldus on eelkõige kohane relatsioonilise andmebaasi väljade kirjeldamisel.

Standardis on kehtestatud kirjelduste „vajalik miinimum“ RIHAKEse rakenduse vaatest – ehk kirjelduselemendid mida on vaja andmeelemendi sisuliseks identifitseerimiseks, omavaheliseks seostamiseks ning andmesõnastikuga sidumiseks. Asutustes hallatav andmeelementide kirjeldus võib olla laiem, sh sisaldada põhjalikumat teavet andmeelemendi tehnilise realisatsiooni kohta (kordsus, kohustuslikkus, vaikeväärtused, täitmise reeglid jne).

| # | Elemendi nimetus | Määratlus ja kasutamine | Kohustuslik / korduv | Näide |
| --- | --- | --- | --- | --- |
| 1 | **tähis** | Andmeelemendi tehniline tähis andmebaasis (välja nimi). Tähis võib olla täheline, numbriline, muu lühend või akronüüm. Tähis on eelistatult semantiliselt arusaadav, kuid ei pruugi seda olla.**Märkus:** Andmeelemendi tähis loetakse RIHAKEse poolt automaatselt andmebaasi skanneerimise käigus. | 1..1 | algus\_kpvhariduseluk\_EHAKjt28 |
| 2 | GUID | Andmeelemendi globaalselt unikaalne identifikaator.**Märkus:** Andmeelemendi GUID luuakse automaatselt RIHAKEse rakenduse poolt.  | 0..1 | 123e4567-e89b-12d3-a456-426655440000  |
| 3 | URI | Nimest, aadressist või tähisest koosnev URI, mis viitab andmeelemendile ainuliselt.**Märkus:** Andmeelemendi URI luuakse automaatselt RIHAKEse rakenduse poolt. | 0..1 | http://rihake/70001234/ebis/andmehoidla2/schema1/table13/section |
| 4 | kirjeldus  | Lühike ja inimloetav andmeelemendi sisu ja tähenduse kirjeldus. **Märkus:** Andmeelemendi kirjeldus loetakse RIHAKEse rakenduse poolt automaatselt andmebaasi skanneerimise käigus (kui on andmebaasis olemas). RIHAKEse rakenduses saab kirjeldust lisada.  | 0..1 | Püsielukoht: elukoht, kus isik veedab enamiku oma igapäevasest puhke- ja uneajast.  |
| 5 | märksõna | Andmeelemendiga seotud andmesõnastiku märksõna termin inimloetaval kujul. | 0..1 | püsielukohtmerihoone |
| 6 | märksõna URI | Viide andmesõnastiku terminile / sõnale juhul kui andmeelement vastab andmesõnastiku terminile või on osaks sellest.**Märkus:** Kirjelduselement täidetakse RIHAKEse rakenduse poolt automaatselt viitega RIHAKEses oleva andmesõnastiku terminile.  | 0..1 | http://rihake/70001234/andmes6nastik/9beff9e8-a26c-456c-a1a1-824b70ec49c9 |
| 8 | seotud loend  | Andmeelemendi täitmisel kasutatud koodiloendi või klassifikaatori viide. **Märkus:** RIHAKEse kasutajaliideses saab kasutaja valida süsteemi sisestatud loendite hulgast sobiva, RIHAKEse sisene viide klassifikaatorile sisestatakse automaatselt.  | 0..1 | http://rihake/70001234/classifiers |
| 9 | **andmetüüp** | Andmeelemendi andmetüüp algses andmebaasis**Märkus:** Andmeelemendi andmetüüp loetakse RIHAKEse poolt automaatselt andmebaasi skanneerimise käigus. | 1..1 | int4(10)varchar(255)uuid |
| 10 | staatus | Andmeelemendi kirjeldatavuse ja kasutatavuse staatus. Staatusel on kolm võimalikku väärtust: „KIRJELDATAV“, „EI KIRJELDATA“, „EI OLE KASUTUSEL“. Vaikimisi väärtuseks on „KIRJELDATAV“ | 0..1 | KIRJELDATAVEI KIRJELDATAEI OLE KASUTUSEL |
| 11 | kirjelduse muutmise aeg | Andmeelemendi kirjelduse viimase muutmise kuupäev ja kellaaeg. Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP)**Märkus:** Kirjelduselement täidetakse RIHAKEse rakenduse poolt automaatselt. | 0..1 |  2022-03-17T10:10:34,9344 |
| 12 | on primaarvõti | Kirjelduselement näitab, kas kirjeldatav andmeelement on andmebaasi tabeli primaarvõti. **Märkus:** Andmeelemendi primaarvõtme staatus loetakse RIHAKEse poolt automaatselt andmebaasi skanneerimise käigus. | 0..1 | falsetrue |
| 13 | on välisvõti | Kirjelduselement esitab viite andmeelemendiga seotud primaarvõtmele.  | 0..1 | riha.comment\_id.comment |

## Sõnastikud

Andmekirjelduse standard esitab kirjelduselemendid andme- ja ärisõnastiku ning nendes sisalduvate terminite kirjeldamiseks.

Andme- ja ärisõnastiku kirjeldus kattub kirjelduselementide mõttes kuid on erinev sõnastike olemuse vaatest – ärisõnastikus sisalduvad iseseisvalt defineeritud, organisatsioonis kasutatavad terminid, andmesõnastik katab reaalselt andmestikus esinevad olulised mõisted. Seega ei pruugi ärisõnastik ja andmesõnastik olla täielikult kattuvad, ärisõnastikus võib eksisteerida termineid mis pole kasutusel andmesõnastikus ja vastupidi.

Andme- ja ärisõnastiku terminite kirjeldus erineb osaliselt ka kirjelduselementide mõttes, eelkõige võimaldab ärisõnastik defineerida terminite hierarhiat, eelistermineid ning esitada ingliskeelse termini ning määratluse.

### Andme- või ärisõnastiku kirjeldus

| # | Elemendi nimetus | Määratlus ja kasutamine | Kohustuslik / korduv | Näide |
| --- | --- | --- | --- | --- |
| 1 | **nimetus** | Sõnastiku nimetus, üldjuhul koos viitega sõnastiku omanik-organisatsioonile. | 1..1 | Kultuuriministeeriumi andmesõnastik |
| 2 | Kirjeldus | Sõnastiku inimloetav kirjeldus, mis viitab selle kasutusvaldkonnale või ulatusele. | 1..1 | Spordi-, loomemajanduse- ja etenduskunstide valdkondade andmestike andmesõnastik |
| 3 | URI | Masinloetav sõnastiku viide URI vormingus. **Märkus:** SõnastikuURI täidetakse automaatselt RIHAKEse poolt, viide võetakse andmesõnastike jaoks kasutusele koos RIHAKEse juurutamisega. | 0..1 | http://rihake/70006317/BV/12c6725b-3778-a5ab-7445-2635d99ea4e7 |
| 4 | **omanik** | Sõnastiku omanik-organisatsiooni nimi. | 1..n | Majandus- ja Kommunikatsiooniministeerium |
| 5 | seotud sõnastikud | Viited sõnastikus kasutatavatele teistele sõnastikele, andmesõnastiku puhul näiteks organisatsiooni ärisõnastikule. Viide esitatakse URI vormingus.**Märkus:** RIHAKEse rakendus võimaldab seotud sõnastiku valimist kasutajaliideses, URI täidetakse rakenduse poolt automaatselt. | 0..n | http://rihake/70006317/BV/12c6725b-3778-a5ab-7445-2635d99ea4e7 |
| 7 | **kasutusele võtmise kuupäev** | Sõnastiku kasutusse võtmise kuupäev.Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP) | 1..1 | 2022-09-23 |
| 8 | viimase muutmise kuupäev | Sõnastiku viimase täiendamise kuupäev.Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP) | 0..1 | 2022-10-19 |

### Andmesõnastiku termini kirjeldus

| # | Elemendi nimetus | Määratlus ja kasutamine | Kohustuslik / korduv | Näide |
| --- | --- | --- | --- | --- |
| 1 | **termin** | Andmesõnastiku termini inimloetav nimetus. Võimalusel kasutatakse sobiva ärisõnastiku termini nimetust. Sobiva termini puudumisel ärisõnastikus luuakse uus andmesõnastiku termin andmekirjelduse käigus.  | 1..1 | Ehitise suletud netopind |
| 2 | määratlus | Andmesõnastiku termini vabatekstiline selgitus. Võimalusel kasutatakse sobiva ärisõnastiku termini määratlust. Sobiva termini puudumisel ärisõnastikus luuakse uus termini määratlus andmekirjelduse koostamise käigus. | 0..1 | Hoone suletud netopind on kõigi korruste suletud netopindade summa |
| 3 | URI | Termini viide andmesõnastikus. Viide esitatakse URI vormingus. **Märkus:** RIHAKEse rakenduses luuakse igale terminile URI automaatselt. | 0..1 | http://rihake/70006317/DD/89c0607b-2940-4b67-8783-2817c92c2ce4/term1234 |
| 4 | kehtivus | Märge selle kohta, kas termin on veel kasutusel (true/false). Termin võib olla kasutusest võetud mitmesugustel põhjustel, näiteks andmekoosseisu muutuse tagajärjel.**Märkus:** Kui element ei ole täidetud, eeldatakse et termin on kasutusel (väärtus: true).  | 0..1 | truefalse |
| 5 | loomiskuupäev | Kuupäev, mil termini kirjeldus lisati andmesõnastikku. Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP). | 0..1 | 2022-02-16 |
| 6 | muutmiskuupäev | Kuupäev, mil termini kirjeldust viimati muudeti. Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP). | 0..1 | 2022-12-11 |
| 7 | seos ärisõnastiku mõistega | Viide terminile ärisõnastikus. Viide esitatakse URI vormingus.**Märkus:** RIHAKEse rakendus võimaldab kasutajaliideses siduda inimloetavaid ärisõnastiku ja andmesõnastiku termineid. URI vormingus viide lisatakse RIHAKEse rakenduses automaatselt.  | 0..1 | http://rihake/70006317/BV/12c6725b-3778-a5ab-7445-2635d99ea4e7/term1234 |
| 8 | märkused | Andmesõnastiku termini vabatekstiline kirjeldus. | 0..1 | Ilmselt muutub haldusreformi tulemusena;Kasutusotstarve ebamäärane |

### Ärisõnastiku termini kirjeldus

| # | Elemendi nimetus | Määratlus ja kasutamine | Kohustuslik / korduv | Näide |
| --- | --- | --- | --- | --- |
| 1 | **termin** | Ärisõnastiku termini inimloetav nimetus  | 1..1 | Projekt |
| 2 | termin (inglise keeles) | Ärisõnastiku termini inimloetav nimetus inglise keeles | 0..1 | Project |
| 3 | **määratlus** | Ärisõnastiku termini vabatekstiline seletus.  | 1..1 | Mingis keskkonnas olulise ja unikaalse muutuse saavutamiseks vajalik ajutine üritus või spetsiifilise juhtimisega ajutine organisatsioon oma kindlaks määratud protsesside ja ressurssidega |
| 4 | määratlus (inglise keeles) | Ärisõnastiku termini vabatekstiline seletus inglise keeles.  | 0..1 | A temporary organisational structure which is setup to create a unique product or service (output) within certain constraints such as time, cost, and quality |
| 5 | termini URI | Termini viide ärisõnastikus. Viide esitatakse URI vormingus. **Märkus:** RIHAKEse rakenduses luuakse URI igale terminile automaatselt. | 0..1 | http://rihake/70006317/BV/12c6725b-3778-a5ab-7445-2635d99ea4e7/term1234 |
| 6 | allikas | Viide termini kirjelduse allikale, näiteks valdkonnasõnastikule, standardile või kirjandusele.  | 0..1 | GSIM versioon 1.2 https://statswiki.unece.org/ display/clickablegsim/Data+Set |
| 7 | on eelistermin | Kirjelduselement näitab, kas termin on sünonüümidega võrreldes eelistatav või mitte. **Märkus:** Kui element ei ole täidetud, eeldatakse et termin ei ole eelistermin (väärtus: false). | 0..1 | truefalse |
| 8 | on kõrgeim tase | Kirjelduselement näitab, kas termin on kõige üldisem antud mõistet kirjeldav termin või mitte. **Märkus:** Kui element ei ole täidetud, eeldatakse et termin on kõrgeim tase (väärtus: true). | 0..1 | truefalse |
| 9 | loomise kuupäev | Kuupäev, mil termini kirjeldus lisati ärisõnastikku.Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP). | 0..1 | 2022-02-16 |
| 10 | muutmise kuupäev | Kuupäev, mil termini kirjeldust viimati muudeti.Kuupäev esitatakse ISO 8601-1 vormingus (AAAA-KK-PP). | 0..1 | 2022-12-11 |
| 11 | märkused | Vabatekstiline kirjeldus termini kohta. | 0..1 | Ilmselt muutub haldusreformi tulemusena;Kasutusotstarve ebamäärane |

1. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe> [↑](#footnote-ref-1)
2. Käesoleva standardi andmeelementide kirjelduse osa on piiratud ainult relatsiooniliste andmebaasidega. Edaspidi on võimalik standardi laiendamine ka teistes vormingutes loodavate ja hoitavate andmestike sisu kirjeldamiseks. [↑](#footnote-ref-2)
3. RIHAKE on rakendus, milles organisatsioonid saavad luua andmestiku üldise kirjelduse, skaneerida relatsioonilise andmebaasina hallatava andmestiku füüsilist mudelit, tuvastada tabelite ja veergude (andmeväljade) olemasolevad metaandmed ning rikastada neid sisulise tähendusega, ehk siduda organisatsioonis kasutatavate mõistetega. [↑](#footnote-ref-3)
4. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-7> [↑](#footnote-ref-4)
5. Alamelementide 12.1, 12.2, 12.3 ja 12.4 kohustuslikkus ja korduvus kehtivad ühe elemendi (kontaktisik) kontekstis. Näiteks võib igas kontaktisiku kirjes olla täpselt üks nimi, samas ei pruugi andmestiku kirjelduses kontaktisikuid olla üldse kirjeldatud (kontaktisiku element ei ole kohutuslik) või on kirjeldatud mitu kontaktisikut (kontaktisiku element on korduv). [↑](#footnote-ref-5)
6. <https://op.europa.eu/en/web/eu-vocabularies/dataset/-/resource?uri=http://publications.europa.eu/resource/dataset/frequency> [↑](#footnote-ref-6)